Characterization of new endogenous geminiviral elements in yam (*Dioscorea* spp.) genomes

Filloux D.[1], Koohapitagtam M.[2], Golden M.[3], Julian C.[1], Galzi S.[1], Rodier-Goud M.[4], D'Hont A.[4], Vernerey MS.[1], Wilkin P.[5], Peterschmitt M.[1], Martin DP.[3], Roumagnac P.[1]

[1] CIRAD-INRA-SupAgro. UMR BGPI, TA A-54/K, Campus International de Baillarguet, F-34398 Montpellier Cedex 5, France. Email: filloux@cirad.fr
[2] Department of Pest Management, Faculty of Natural Resources, Prince of Songkla University, Hat Yai campus, Thailand 90120
[3] Computational Biology Group, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town 4579, South Africa
[4] CIRAD, UMR DAP, TA A-96/03, Avenue Agropolis, F-34398 Montpellier Cedex 5, France
[5] Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AB, U.K.

filloux@cirad.fr

Geminivirus sequences integrated into the genomes of various plant species (*Nicotiana* species, *Lactuca sativa*, *Malus domestica* and *Populus trichocarpa*) have been reported. The best studied of these integrated geminivirus sequences are the so-called "geminivirus-related DNA" or GRD elements within the genomes of various *Nicotiana* species. Our study aimed at assessing whether such integrated geminiviral sequences are present within the genomes of yam species. We have used a polyphasic approach combining a first step of plant genome screening (*in silico* EST analysis and viral metagenomics) followed by a second step of molecular characterization of the integrated sequences (Long-PCR and Inverse-PCR). We have found two new phylogenetically related endogeneous geminivirus (EGV) elements within the genome of the majority of Asian *Dioscorea* spp. section *Enantiophyllum*. Based on multiple lines of evidence, such as (i) seed transmission of the two EGV sequences, (ii) no amplification by rolling circle amplification, (iii) worldwide presence of the two EGV sequences in a range of asympomatic Asian *Dioscorea* species and (iv) lack of sequences of coat protein genes (i.e. they are not likely to be insect transmissible), we hypothesized that those two EGV elements were likely to have become integrated within an ancestor of the Asian *Dioscorea* species. We then confirmed the presence of the two EGV elements within *D. alata* and *D. nummularia* genomes by fluorescence *in situ* hybridization. The predicted proteins expressed by the two EGV sequences had detectable homology to begomovirus replication enhancer and replication associated proteins but only one out of the two EGV sequences contains a GC rich sequence resembling the conserved hairpin structures found at geminivirus virion strand origins of replication (including the TAATATTAC sequence in the likely loop region) and is organized as 2.6 Kb tandem repeats. The discovery and analysis of such fossilized geminivirus sequences within plant genomes would certainly help to more accurately date events deep in the evolutionary history of the *Geminiviridae* family such as the most recent geminivirus common ancestor, and the origins of the various geminivirus genera.

Mots-clés: Integrated viral sequences, Geminivirus, *Dioscoreacae*, *Enantiophyllum*, Phylogeny, FISH